# Compatible and incompatible representations in visual sensory storage

**Rishi Bhardwaj**
Department of Psychology, Durham University, Durham, UK

**John D. Mollon**
Department of Experimental Psychology, University of Cambridge, Cambridge, UK

**Hannah E. Smithson**
Department of Psychology, Durham University, Durham, UK
Department of Experimental Psychology, University of Oxford, Oxford, UK

Sensory storage shows a short-lived part-report advantage that survives an aftercoming visual noise pattern (Smithson & Mollon, 2006). We tested whether such an advantage survives different types of high-contrast mask. The target was a 3 × 4 array of digits. The mask could be (a) a noise pattern, (b) an array of eights, or (c) an array of random digits. In a preliminary experiment, target and mask were interleaved (at 140 Hz) and target contrast was varied to determine the level at which performance fell to chance. In the main experiment, target and mask were separated by an inter-stimulus-interval (ISI) of 100, 150, or 200 ms. An auditory part-report cue that was presented 240 ms after target offset supported a part-report advantage at all ISIs for noise masks, at ISIs greater than 100 ms for digit-8 masks, but not at any ISI for random-number masks. Increasing cue delay, in the range 240 to 730 ms, produced a decline in the advantages we measured. The differences in part-report superiority with different types of mask call for revision of the model of visual sensory storage as a single canvas on which successive items are superposed. When mask and target are sufficiently different, a representation of low-contrast target digits can be maintained independently of the representation of an aftercoming, high-contrast mask. However, when the same target is followed by a mask composed of high-contrast random digits, an independent representation of the target does not remain available for access.

Keywords: iconic memory, visual short-term memory, part-report, pattern masking, backward masking

Citation: Bhardwaj, R., Mollon, J. D., Smithson, H. E. (2012). Compatible and incompatible representations in visual sensory storage. *Journal of Vision, 12*(5):1, 1–10 http://www.journalofvision.org/content/12/5/1, doi:10.1167/12.5.1

## Introduction

The visual world delivers a rich and continuous flow of information to our eyes, and we must select a subset of this information to guide our behavior. The relevance of particular information is not always known at the moment of presentation, and the brain is thought to maintain a sensory buffer so that inputs can be briefly held for further analysis if they prove relevant even after the physical source has disappeared. The brief sensory memory was termed the icon by Neisser (1967) and was classically demonstrated by the part-report technique (Averbach & Coriell, 1961; Averbach & Sperling, 1961; Sperling, 1960; Sperling, 1983). Iconic storage has been differentiated from visual short-term memory (VSTM): the latter is of smaller capacity, lasts for many seconds, and is not susceptible to masking (Phillips, 1974). However, it has become difficult to distinguish iconic storage from a fragile form of VSTM since a part-report effect and susceptibility to backward masking are common to both (Sligte, Scholte, & Lamme, 2008; Sligte, Vandenbroucke, Scholte, & Lamme, 2010; for discussion see Smith, Mollon, Bhardwaj, & Smithson, 2011).

In many branches of cognitive science, researchers use backward masking to limit the duration for which a stimulus is available for processing (e.g., Van Orden, 1987; Rayner, Smith, Malcolm, & Henderson, 2009): the aftercoming mask is held to degrade or destroy the icon (or "fragile VSTM") and thus make further visual encoding impossible. The rationale is that if a stimulus is presented for, say, 50 ms then the stimulus enters the iconic store and attention can later be directed to it. However, if the stimulus is followed by a mask then what remains in the iconic store is either a mixed representation of the stimulus and mask or only the mask. The use of masks to "terminate processing" has survived (Reeves, 2007) despite the warnings made by Charles Eriksen three decades ago (Eriksen, 1980;

doi: 10.1167/12.5.1 Received August 20, 2011; published May 4, 2012 

Schultz & Eriksen, 1977). Recently, Smithson and Mollon (2006) provided evidence for continued access to the target even after a mask has been presented. They showed that a part-report advantage can be obtained even when the part-report cue follows a potent pattern mask. They presented their observers with an array of letter stimuli followed first by the pattern mask and then by a visual cue indicating which row of letters to report. The part-report advantage was greatest when the cue immediately followed the mask and was reduced when the cue was delayed by 500 ms. The latter result suggested that the part-report advantage depended on a very short-lived store.

We here test the generality of Smithson and Mollon's result. We ask whether an analogous part-report effect can be found for different types of mask: Can the representation of the target survive in the sensory buffers, whatever the type of mask that follows it, or are some target and mask combinations incompatible? Our target is always an array of 12 randomly chosen digits. One mask is a random-pattern mask as before. In a second mask, a digit eight is presented in each of the 12 positions of the array and in the same font as the target. A third mask is an array of random digits. We use three different inter-stimulus-intervals (ISIs) between the target and the mask (100, 150, and 200 ms). In the case of all these ISIs, the target is visible to the participant. Our interest is not in the absolute effect of the mask but in the relative performance for part-report compared with whole-report. This difference is a measure of whether a short-lived representation of the target survives the onset and offset of the high-contrast mask.

It is important for our purpose to show that the report of the target array is at chance if target and mask are combined. It would be trivial to show survival of part-report superiority for a relatively weak mask that allowed recovery of target information from a sensory representation in which mask and target were integrated. By testing performance when the mask and target are interleaved on alternate frames we are able to choose a target contrast that is sufficiently low that recovery of target information from an integrated representation is at chance.

We show that part-report effects as large as 32–43% can be found when the part-report cue follows a pattern mask. A smaller but significant part-report effect is found with digit-8 masks, but there are no reliable effects when the mask is formed from random digits.

# Methods

## Observers

Four observers (three males, RJL, YK, RB, and one female, HES) with a mean age of 30 participated in this experiment after giving their informed consent. All observers had corrected-to-normal visual acuity. RJL and YK were naïve as to the aims of the experiment and were paid for their participation.

## Apparatus

The experiment was implemented by a custom-written Matlab program. Visual stimuli were presented on a 20-inch Mitsubishi Diamond Pro 2070SB color monitor driven by a Cambridge Research System (CRS) ViSaGe graphics system. The monitor's frame rate was 140 Hz and the screen resolution was 1024 × 769 pixels. The monitor screen was gamma corrected using a CRS OptiCAL photometer. The experiment was performed in a dark room so that the only significant source of light came from the monitor screen. Auditory stimuli were presented by Labtec speakers driven by a Realtek onboard soundcard. Physical synchronization of the auditory and visual stimuli was checked empirically using an oscilloscope to compare the signal from the soundcard and the output from a fast photodiode. Observers responded using a wireless number-pad.

## Stimuli

The background luminance of the stimulus monitor was 15.70 cd/m$^2$; target and mask stimuli were increments from this level and had a luminance of 30.85 cd/m$^2$. Target stimuli were 3 × 4 arrays of digits generated by selecting digits at random from the set of zero to nine with replacement. The performance level for correctly guessing a digit from an array was therefore one out of ten. The digits were displayed using a seven-segment "DS-Digital" font. The height of each letter subtended 0.46° of visual angle at the viewing distance of 1.3 m. Each of the 12 digits in the target array was separated horizontally by 0.41° and vertically by 0.41°. The entire target array subtended 2.44° horizontally and 2.29° vertically. Fixation was guided by a low-contrast cross which was present at the start of each trial.

Mask stimuli could be (a) a 4.12° × 4.12° patch of binary visual noise with element size equal to the stroke width of the target digits (0.06°), or (b) a 3 × 4 array of the digit eight, or (c) a 3 × 4 array of random digits. The font and the spatial arrangement of the digit-8 mask and of the random-digit masks were identical to those of the targets. As in the case of the targets, the random-digit masks were generated by drawing randomly from the set of zero to nine with replacement.

At the beginning of each trial, two 877-Hz, 20-ms warning tones were presented with an interval of 1

second between them. The target array was presented 1 second later. In part-report conditions, the 20-ms auditory cue could be a high tone (1.7 kHz) indicating that participants should report the digits from the first row, a medium tone (877 Hz) indicating the second row, or a low tone (434 Hz) indicating the third row. The warning tone acted as a reference i.e., the cue-tone was higher than, the same as, or lower than the warning tone.

## Procedure

### Preliminary experiment

A preliminary experiment was conducted to ensure that the different types of mask were matched in their ability to obscure the target in a combined representation. In a short train of six frames (which, at the adapting luminance, fell within Bloch's law for total temporal integration of contrast energy, Gorea & Tyler, 1986), target and mask were presented on alternate frames. Mask contrast was fixed at 100% but target contrast was varied in the range 25–64% according to a method of constants to establish the maximum contrast that gave chance performance. The row to be reported was cued, either with a precue presented 729 ms before the onset of the stimulus train or with a postcue presented 100 ms after the offset of the stimulus train. The two types of cue and the three types of mask were tested in separate counterbalanced blocks, giving six blocks in one experimental session. Within each block, there were 75 randomized trials (5 contrast levels × 3 cued rows × 5 repeats). Each observer participated in seven sessions. The first session was discarded as practice. Observers were instructed to report the target digits in their correct spatial position and to guess any digits about which they were unsure.

On the basis of these preliminary measurements, the contrasts chosen for the main experiment were 32% for HES, YK, and RB, and 40% for RJL. At these values, performance for each observer did not differ significantly from chance for any mask type.

### Main experiment

In the main experiment, we used three types of mask and three values, 100, 150, or 200 ms, for the interstimulus interval (ISI) between the offset of the target and the onset of the mask. There were three kinds of experimental block: part-report with precues, part-report with postcues, and whole-report. In precue part-report blocks, the cue was presented 729 ms before the onset of the target; there were 15 trials (3 cued rows × 5 repeats). In the postcue part-report blocks, the cue was presented 0, 121, 243, 364, 486, 607, or 729 ms after the target offset and there were 105 trials (7 cue delays × 3 cued rows × 5 repeats). In whole-report blocks there were five repeats. In each experimental session we tested each type of block for each type of mask but one ISI; different ISIs were tested in different sessions. All conditions were repeated seven times, with the first being discarded as practice.

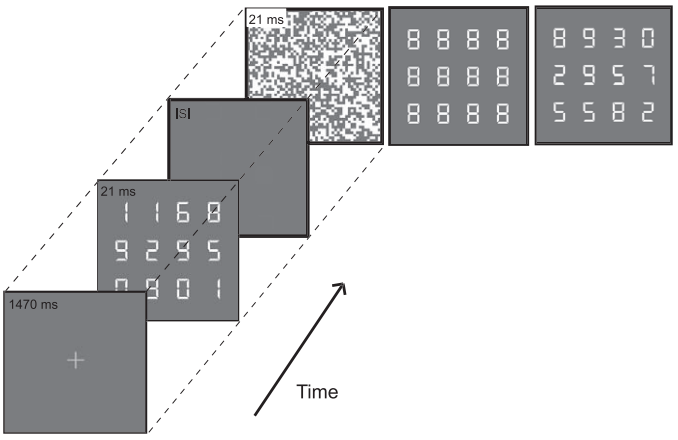


Figure 1. Schematic representation of the stimuli in the main experiment. The target comprised a 3 × 4 array of digits presented for three consecutive frames (at a frame-rate of 140 Hz, giving a nominal duration of 21 ms). The target was followed by a uniform field for (a) 14 frames (100 ms), (b) 21 frames (150 ms), or (c) 28 frames (200 ms) and then by the mask which could be (a) a binary noise mask, (b) a 3 × 4 array of digit-8s, or (c) a 3 × 4 array of digits, presented for 3 frames (21 ms). For the part-report conditions, an auditory cue was presented (with a delay of −750, 0, 121, 243, 364, 486, 607, or 729 ms relative to the target offset) to instruct participants to report the numbers from the top, middle, or bottom array.

The sequence of visual stimuli was similar to that of the preliminary experiment except that target and mask were now separated in time (Figure 1). Following two warning tones, the target array, set at the contrast chosen on the basis of the preliminary experiment, was presented for three frames (nominal duration of 21 ms). It was followed by a uniform field for the appropriate ISI: 14 frames (100 ms), 21 frames (150 ms), or 28 frames (200 ms). The mask was then presented at 100% contrast for three frames (21 ms).

As in the preliminary experiment, auditory cues were used to indicate which row of digits to report in part-report trials. In the whole-report blocks no auditory cues were presented and the observers had to report all three rows. As in the preliminary experiment, observers were asked to report the target digits in their correct spatial position and to guess any digits about which they were unsure.

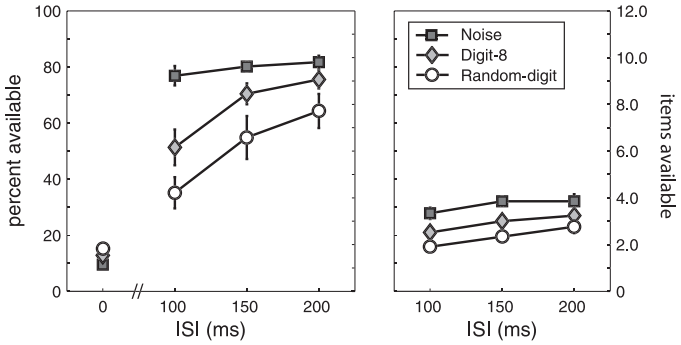


Figure 2. Average data collected from four participants for the precue (left-hand panel) and whole-report conditions (right-hand panel). The *x*-axis represents the time between the offset of the target and the onset of the mask (ISI). Plotted at ISI = 0 are data from the preliminary experiment. Different symbols are used for different mask-types: binary-noise (dark grey squares), digit-8s (light grey diamonds), and random-digits (white circles). The *y*-axis represents the estimated number of digits available to the participant, expressed as a percentage on the left hand side, and as number of items on the right hand side. For the whole-report conditions, the estimated number of digits available is simply the number of digits correctly reported. For part-report conditions it is the number of digits reported per trial, multiplied by three (given that there were three possible cued rows). The error bars associated with each symbol are (+/– 1 *SE*) across participants.

# Results

A correct response was recorded only when participants reported the correct digits in the locations in which they had occurred in the target array. In the whole-report conditions the estimated number of digits available is the sum of correctly reported digits across all rows. In part-report conditions it is the number of digits reported per cued row, multiplied by three (given that there were three rows, cued at random). Data were analyzed with repeated-measures analyses of variance (ANOVAs). In all cases reported below, Mauchly's test indicated that sphericity could be assumed.

## Precue condition

The precue condition allows us to estimate the visibility of the target arrays when the observer knows in advance the row to be reported. The results for this condition are shown in the left-hand panel of Figure 2, where the leftmost data points indicate performance when target and mask are interleaved in the preliminary experiment. When an ISI of 100 ms or more is present, the targets are above threshold for all masking conditions, but the three types of mask differ clearly in their effect on performance: the binary-noise masks have the smallest effect and the random-digit masks have the greatest. A two-way ANOVA showed that the effects of ISI and of mask types were highly significant ($F(2, 6) = 145.826$, $MSE = 16.908$, $p = 0.0001$; $F(2, 6) = 34.147$, $MSE = 36.405$, $p = 0.001$). There was also a significant interaction ($F(4, 12) = 12.120$, $MSE = 2.430$, $p = 0.001$), although this may reflect ceiling effects for noise masks. A subsidiary analysis showed that performance did not differ according to cued row.

## Whole-report condition

A similar difference between mask types is seen in the whole-report condition (Figure 2, right-hand panel). A two-way ANOVA showed that the effects of ISI and of mask type were highly significant ($F(2, 6) = 15.387$, $MSE = 1.383$, $p = 0.004$; $F(2, 6) = 134.291$, $MSE = 5.636$, $p = 0.0001$). There was no significant interaction.

## Part-report effects in the postcue condition

Figure 3 shows performance as a function of cue delay in the postcue and part-report conditions. Each row of panels corresponds to a different ISI and each column to a different type of mask. Whole-report performance is indicated by the horizontal line in each panel.

In assessing the data of Figure 3, we ask—for each type of mask and each ISI—whether there is a significant part-report advantage and whether this advantage declines with cue delay. These are features consistent with a brief sensory memory of the targets that participants can still access after the onset and offset of the high-contrast mask. Such features are present in the first column, which represents results for the binary-noise masks, and they are clearly absent in the third column, which gives results for the random-digit masks. The results for the digit-8 mask, shown in the middle column, are intermediate.

A 3 × 3 × 7 repeated-measures ANOVA was performed on the estimated number of digits available in all the postcue conditions, with ISI, mask-type, and cue-delay as factors. As in the case of the precue conditions, performance significantly increased with increasing ISI ($F(2, 6) = 156.628$, $MSE = 24.7$, $p = 0.0001$). Performance was better for the binary-noise masks followed by the digit-8 mask and in turn followed by random-digit masks ($F(2, 6) = 264.802$, $MSE = 111.405$, $p = 0.0001$). Performance was better with early cues than with late cues ($F(6, 18) = 9.02$, $MSE = 1.226$, $p = 0.0001$). There were significant interactions between ISI and mask-type ($F(4, 12) = 9.632$, $MSE = 0.672$, $p = 0.001$) and between cue-delay

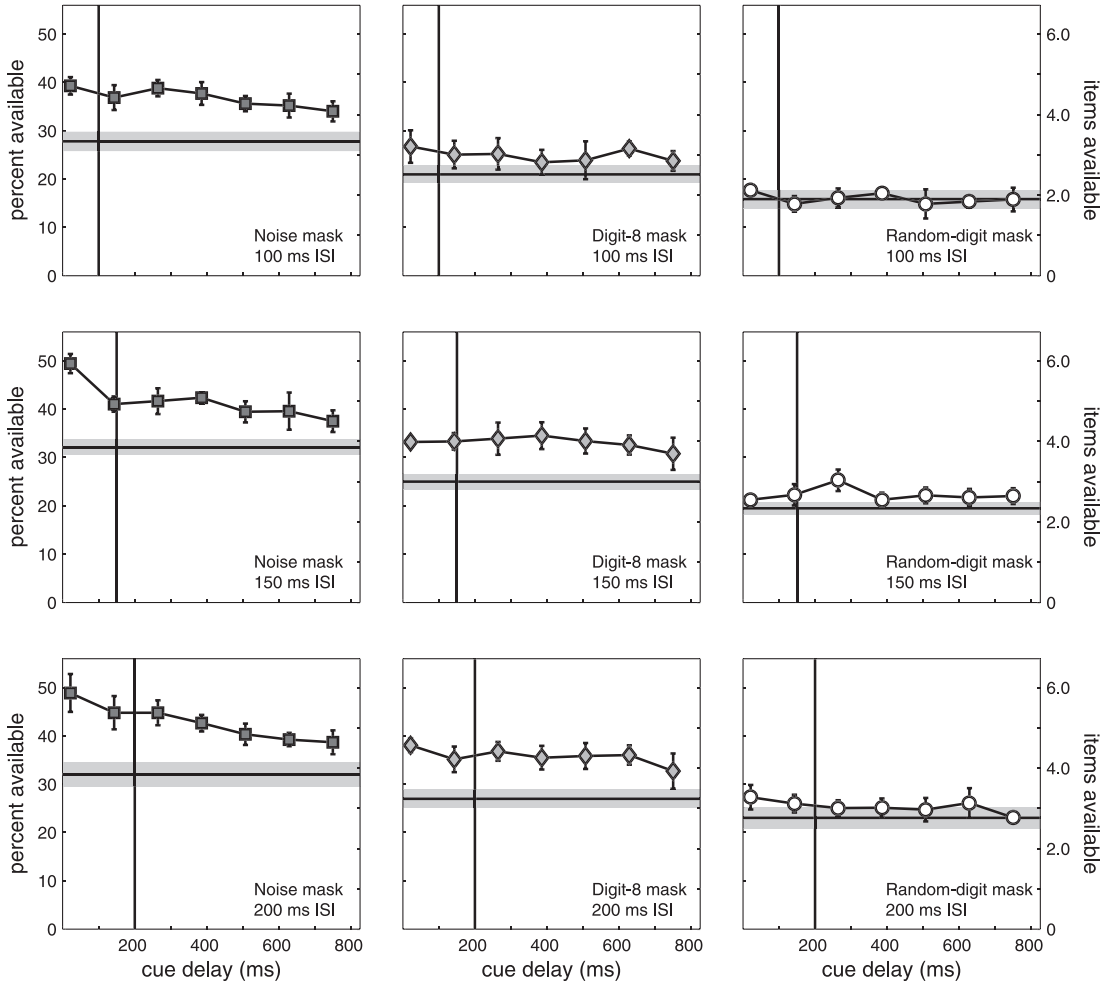


Figure 3. Average data collected from four participants in the main experiment. The columns represent data for different types of mask: binary-noise (left); digit-8s (center); random-digits (right). The rows represent data for different ISIs: 100 ms (top); 150 ms (middle); 200 ms (bottom). The vertical line on each graph shows the temporal position of the mask relative to the offset of the target. The symbols represent data from the part-report conditions in which the delay between the offset of the target and the auditory cue was 0, 121, 243, 364, 486, 607, or 729 ms. The horizontal line represents performance for the whole-report condition. The *y*-axis represents the estimated number of digits available to the participant, expressed as a percentage on the left hand side and as number of items on the right hand side. For the whole-report conditions, the estimated number of items available is the number correctly reported per trial. For part-report conditions it is the number of digits reported per trial, multiplied by three (given that there were three possible cued rows). The error bars associated with each symbol, and the greyed region associated with the horizontal line, represent +/− 1 *SE* across participants.

and mask-type ($F(12, 36) = 4.301$, $MSE = 0.332$, $p = 0.0001$).

For each condition we tested for a part-report advantage that survives masking using a paired samples *t*-test to compare whole-report performance and part-report performance following a postmask cue. We selected data for a cue-delay of 242 ms relative to target offset. This is the condition in which the cue followed the mask even at the longest ISI. Since later cues are predicted to produce smaller part-report effects this is a conservative choice. The results of the analyses are summarized in Table 1. Applying the Bonferroni correction for the three comparisons within each mask-type ($p_{crit} = 0.05/3 = 0.017$), we have robust evidence for significant part-report advantages for noise masks presented at 150 and 200 ms ISI and for the digit-8 mask presented at 200 ms ISI, and marginal evidence for advantages for noise masks presented at 100 ms ISI and for the digit-8 mask presented at 150 ms ISI.

To evaluate decline in performance with increasing cue-delay we conducted a one-factor ANOVA for each mask-type at each ISI. We used only the five cue-delays that caused the cue to follow the mask at all target-mask ISIs. The results of the analyses are summarized in Table 2. For noise masks there is evidence of a significant decline with increasing cue-delay at all ISIs, although the *p*-values are marginal at 150 and 200 ms

| | Binary-noise | | Digit-8s | | Random-digits | |
|---|---|---|---|---|---|---|
| ISI (ms) | $t(3)$ | $p$ | $t(3)$ | $p$ | $t(3)$ | $p$ |
| 100 | 4.616 | 0.019 | 1.939 | 0.148 | 0.103 | 0.924 |
| 150 | 7.305 | 0.005* | 4.629 | 0.019 | 3.180 | 0.050 |
| 200 | 5.968 | 0.009* | 13.012 | 0.001* | 0.845 | 0.460 |

Table 1. Paired $t$-tests comparing part- and whole-report performance (scored including position) with post-mask cue-delays. ISI = inter-stimulus-interval.

ISI. For the digit-8 mask there is evidence of a significant decline with increasing cue-delay only at 150 ms ISI. For random-digit masks there is no effect of cue-delay on part-report performance.

It is possible that the decline in performance with cue-delay results from a loss of spatial information, while identity information remains preserved (e.g., Mewhort, Marchetti, Gurnsey, & Campbell, 1984). We repeated the analysis but scored a response correct if that item occurred anywhere within a row. This analysis gave higher scores in all conditions (as expected from higher chance performance of 1.21 letters out of four instead of 0.4 letters out of four) but the overall pattern of results (part-report advantages that decline with increasing cue-delay) and differences between masking conditions were broadly preserved. The corresponding statistics are summarized in Tables 3 and 4. This result implies that the differences between conditions lie in the loss of identity information and not simply in the mis-binding of items to particular locations.

# Discussion

## Challenging the textbook account of the icon

In the experiments reported here we use a part-report postcue that might allow selective retrieval from a store of shorter duration and higher capacity than that which supports performance in whole-report conditions. Several previous studies (Kahneman, 1968; Turvey, 1973) have compared the effect of masks with different visual characteristics, using reduction in *overall* performance as the metric. We focus specifically on the reduction in *part-report* performance that is produced by different masks. Our reasoning is this: A late-coming cue can support a part-report advantage only if it allows selective report from a representation that preserves information about the target. If, at the time of presentation of the cue, the iconic representation of the target is unavailable (either because it has decayed or because it has been replaced by, or integrated with, a destructive mask) performance with the cue will be in direct proportion to whole-report performance and will show no advantage over performance with a cue that is further delayed. In these cases, performance will depend on what can be extracted before the mask is presented and on anything that can be retrieved from a combined representation of mask and target. Conversely, any part-report advantage that we measure indexes a representation of the target that remains separate from the mask.

Our results are not compatible with an account in which the icon is a single static canvas on which successive items are superposed. This conclusion is in part anticipated by the dissociation between the effects of the masks at zero ISI (when target and mask are temporally interleaved) and at long ISI. Stimulus contrasts that were chosen to produce equal performance for all types of mask at zero ISI produce different levels of performance for different mask types at long ISIs (Figure 2). This echoes the classical distinction between integration and interruption forms of masking, which dominate at ISIs around zero and 100 ms respectively, and for which different mask characteristics are important (Enns and Di Lollo, 2000). At each ISI, the different mask conditions are indistinguishable until presentation of the mask.

| | Binary-noise | | | Digit-8s | | | Random-digits | | |
|---|---|---|---|---|---|---|---|---|---|
| ISI (ms) | $F(1, 3)$ | $MSE$ | $p$ | $F(1, 3)$ | $MSE$ | $p$ | $F(1, 3)$ | $MSE$ | $p$ |
| 100 | 15.714 | 0.841 | 0.029* | 0.0 | 0.0 | 0.987 | 0.674 | 0.034 | 0.472 |
| 150 | 9.057 | 0.711 | 0.057 | 26.042 | 0.387 | 0.015* | 4.272 | 0.210 | 0.131 |
| 200 | 9.259 | 1.406 | 0.056 | 1.829 | 0.336 | 0.269 | 1.741 | 0.054 | 0.279 |

Table 2. Linear-trend analyses for part-report performance (scored including position) as a function of cue-delay with post-mask cue-delays. ISI = inter-stimulus-interval.

| | Binary-noise | | Digit-8s | | Random-digits | |
|---|---|---|---|---|---|---|
| ISI (ms) | $t(3)$ | $p$ | $t(3)$ | $p$ | $t(3)$ | $p$ |
| 100 | 3.305 | 0.046 | 1.653 | 0.197 | 0.987 | 0.396 |
| 150 | 12.075 | 0.001* | 5.195 | 0.014* | 1.667 | 0.194 |
| 200 | 14.522 | 0.001* | 4.957 | 0.016* | 0.853 | 0.456 |

Table 3. Paired *t*-tests comparing part- and whole-report performance (scored ignoring position) with post-mask cue-delays. ISI = inter-stimulus-interval.

Differences between conditions must therefore arise in the *interaction* between mask and target, and not in the amount of information extracted from the target before presentation of the mask. The differences in relative potency of different masks at short and long ISI suggest that the nature of the target-mask interaction must differ from simple superposition of successive items.

But the crux of our argument depends upon the part-report postcue data. For pattern masks, there is a clear part-report effect when the cue follows the high-contrast mask. If successive items were superposed in the icon, then it should be impossible to recover the low-contrast target from the single integrated representation.

## Why do pattern masks and random-digit masks have different effects?

Our part-report data suggest that, within short-lived visual storage, the interaction between a target and an after-coming mask depends on the similarity between them. Setting aside for the moment the intermediate case of digit-8 masks, we first compare performance with the most different masks: pattern masks and random-digit masks. The representations of digit targets and pattern masks seem to be able to coexist in short-term storage whereas the representations of digit targets and random-digit masks cannot. We consider two accounts. These accounts are cast in the language of two separate traditions but overlap conceptually.

A) Visual processing is not instantaneous, nor is it confined to a single layer of neural processing. Instead, visual information moves through a bank of analyzers of increasing selectivity. One hypothesis, outlined by Smithson and Mollon (2006), is that the masks differ in how far they can pursue the target through this hierarchy. By this account part-report superiority is maintained for pattern masks because the target occupies a level of analyzers that the mask cannot recruit, where target stimuli are represented in terms of abstract visual features or categorized objects or "proto objects" (e.g., Di Lollo, Enns, & Rensink, 2000). The random-digit mask itself consists of exemplars of targets and so will occupy the very same analyzers. The part-report advantage is then abolished either as a result of degradation of the signal within the analyzers or by the initiation of competing responses. There is a growing body of evidence to suggest that stimulus encoding and storage depend on the same feature-selective neural circuits (Pasternak & Greenlee, 2005). At retention intervals typically associated with VSTM, there is some evidence to suggest that different perceptual attributes are represented in parallel memory stores (Magnussen & Greenlee, 1999). Similarly, in the case of iconic memory, the textbook account of a single snapshot icon should perhaps be replaced by an account in which multiple iconic stores exist, each for a different class of stimulus. These multiple iconic stores are not mysterious: They would correspond to the analyzers for different classes of stimulus.

B) The devastating effect of the random-digit masks is consistent with an account in which short-lived visual memory cannot represent multiple objects at the same spatial position. The idea that incompatible representations must somehow be resolved into a single percept is recurrent in accounts of visual masking (Turvey, 1973). Enns and Di Lollo (2000; Di Lollo et al. 2000) present a model of "reentrant processing" that checks

| | Binary-noise | | | Digit-8s | | | Random-digits | | |
|---|---|---|---|---|---|---|---|---|---|
| ISI (ms) | $F(1, 3)$ | *MSE* | $p$ | $F(1, 3)$ | *MSE* | $p$ | $F(1, 3)$ | *MSE* | $p$ |
| 100 | 6.035 | 0.441 | 0.091 | 0.269 | 0.006 | 0.640 | 0.002 | 0.0002 | 0.969 |
| 150 | 34.687 | 0.361 | 0.010* | 9.050 | 0.240 | 0.057 | 0.925 | 0.078 | 0.407 |
| 200 | 147.682 | 1.444 | 0.001* | 8.339 | 0.633 | 0.063 | 0.710 | 0.136 | 0.461 |

Table 4. Linear-trend analyses for part-report performance (scored ignoring position) as a function of cue-delay with post-mask cue-delays. ISI = inter-stimulus-interval.

rapidly changing incoming information against a more slowly updated perceptual hypothesis. Keysers and Perrett (2002a) propose an alternative "neural competition" account based simply on relative signal strength. The present experiment is not a test of visual masking per se: Our interest is in the use of a late-coming cue to report selectively from a decaying representation of the target digits and the ability of a mask to disrupt this representation. It is possible that at each position in the target array, the iconic representation might hold only one digit, and that the target digit is lost following presentation of a random-digit mask because the sensory evidence is stronger for the digit from the mask. By this account, the pattern-mask fails to displace the digit-objects from the iconic representation: This could be because the entire noise-pattern is treated as a single object at a different spatial scale from the target digits (Enns & Di Lollo, 1997). Alternatively the pattern-mask might be represented in iconic memory as a transparent veil over the target digits. Keysers and Perrett (2002a; 2002b) identify the perception of transparent materials as an exception to the general rule that multiple physical objects cannot project to one retinal location.

## Why is the digit-8 mask intermediate in its effect?

The regular array of elements in this mask is limited to only one of the possible target identities, and always the same one. Thus this mask is *groupable:* the regular elements can be grouped into a single object and—in an object-substitution account (Enns & Di Lollo, 1997) for example—the large single object is less able to displace individual digits, owing to its different spatial scale. The random-digit mask will exercise the form of masking that Watson, Borthwick, and Taylor (1997) have called "entropy masking" since an ideal observer would not be able to subtract such a mask from a composite representation, whereas the digit-8 mask, being presented in blocked trials and thus fully *predictable*, would in principle allow such a subtraction. Finally the *identity* of the elements of the digit-8 mask would allow participants to improve their overall score by applying a selective attenuation, or an elevated criterion, to representations of the digit eight, either at the level of "feature analyzers" or of "object files".

If participants learn selectively to attenuate representations of the digit eight in the case of the digit-8 mask, then we should expect sensitivity to be decreased on those trials where the target is itself an eight, in contrast to trials where the target is not an eight. We might also expect sensitivity for target eights to be lower in the case of digit-8 masks than in the case of random-digit masks. With only a single pair of hit and false alarm rates for each condition, our data do not allow us to test the assumptions for calculation of d-prime (see Macmillan & Creelman, 2005, Figure 3.10). However, within the framework of signal detection theory, conditions that differ only in false alarm rate (or that differ only in hit rate) are special cases that do permit unambiguous identification of the condition with lower sensitivity. An analysis of our data shows that for the digit-8 mask condition the difference between target-8 and target-not-8 is only in $p$(false alarm), with participants showing lower sensitivity for detection of eights, and that the difference between target-8 in the digit-8 mask condition compared to the random-digit mask condition is also only in $p$(false alarm), with the digit-8 mask condition producing lower sensitivity.

## The relationship of the present results to those of Smith et al. (2011)

Using an experimental paradigm distinct from the present one, Smith et al. (2011) demonstrated conditions in which two successive arrays of digits could be concurrently held in a visual sensory memory. The two arrays of four digits were spatially coincident, and each was followed by a pattern mask similar to the pattern masks used in the present experiment. When the stimulus train was followed by a tone that indicated whether the participant should report the first or the second array, a clear part-report advantage was obtained. This result suggested that the participant had independent access to the two successive arrays within one sensory store. Yet when random digits are used as masks in the present experiment, our participants appear unable to store two successive arrays of digits. How should this apparent discrepancy be explained? In the present experiment, the mask was of much higher contrast than the target, as is conventional in masking experiments. In the experiment of Smith et al (2011), the two sets of digits were both presented at high contrast and followed by pattern masks, and precue performance indicated that the two arrays were equally legible. This is likely to be the critical difference between the two paradigms.

## Implications for studies that use masks to "terminate perceptual processing"

In the classical account an aftercoming pattern mask is held to terminate the processing of an earlier brief array—either by triggering a new snapshot or by the superposition of target and mask within the same snapshot—and it is on this assumption that backward

masking is very often used in experimental psychology. A psycholinguist might for example use backward masking to limit the amount of time a stimulus is available for visual processing, arguing that without the mask the stimulus would remain available in iconic memory. We have shown however that even after presentation of a high-contrast pattern mask, there is a brief period in which a late spatial cue can be used to selectively retrieve information about the target, suggesting that a representation of the target remains accessible in a short-lived store. Only with high-contrast random-digit masks do we achieve abolition of the part-report advantage. This presents a problem for experiments on psycholinguistics: Meaningful characters are unsuitable masks in these experiments (since such masks would contaminate the experiment rather than simply terminate processing) and yet we have shown that pattern masks do not terminate processing.

In an influential and widely cited paper Rayner, Inhoff, Morrison, Slowiaczek, and Bertera (1981) use a gaze-contingent masking paradigm to infer the time required to extract from each fixation the visual information necessary for reading. The participant's eye movements were measured during reading and, a variable time after the start of each fixation, a mask replaced the text at fixation. With delays as short as 50 ms, reading became possible. But, if the mask had failed to terminate the perceptual processing of the text, the time required to extract the necessary visual information could in fact be much longer than 50 ms. Indeed, later information could be used to guide selective extraction of information from a sensory store, as the late cue in our experiments supported an advantage in selective reporting of target digits.

Although "snapshot" models of iconic memory have been challenged in several ways (Haber, 1983), they are still pervasive in experimental psychology (e.g., Simons, 2000) and, nearly 30 years on, the use of a mask to terminate processing is still widespread. To give a recent example, Rayner et al. (2009) used a masking paradigm to estimate the time required to encode stimulus properties during scene perception. After a certain time in each fixation they replaced the scene with a contrast-matched color noise mask. In contrast to the reading study above, they found that normal scene perception required a delay of 150 ms from the start of each fixation to the presentation of the mask. However, if presentation of the mask does not terminate perceptual processing, this measure cannot be used as a secure indicator of the time required for stimulus encoding. Differences between estimates for reading and for scene perception may have more to do with target-mask interactions than with differences in the time required to encode the relevant visual information.

## Acknowledgments

This study was supported by ESRC grant RES-062-23-0530. We thank Ron Rensink for his comments on an earlier version of this article.

Commercial relationships: None
Corresponding author: Rishi Bhardwaj
Email: dr.rishi.bhardwaj@gmail.com
Address: Nethradhama School of Optometry, Bangalore, India

## References

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal,* 40(1):309–328.

Averbach, E., & Sperling, G. (1961). Short term storage of information in vision. In C. Cherry (Ed.), *Information Theory* (pp. 196–211). London: Butterworth.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of re-entrant visual processes. *Journal of Experimental Psychology: General,* 129(4):481–507.

Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science,* 8(2):135–139.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences,* 4(9):345–352.

Eriksen, C. W. (1980). The use of a visual mask may seriously confound your experiment. *Perception and Psychophysics,* 28(1):89–92.

Gorea, A., & Tyler, C. W. (1986). New look at Bloch's law for contrast. *Journal of the Optical Society of America A,* 3(1):52–61.

Haber, R. N. (1983). The impending demise of the icon: A critique of the concept of iconic storage in visual information processing. *Behavioral and Brain Sciences,* 6(1):1–11.

Kahneman, D. (1968). Method, findings and theory in studies of visual masking. *Psychological Bulletin,* 70(6):404–425.

Keysers, C., & Perrett, D. I. (2002a). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences,* 6(3):120–125.

Keysers, C., & Perrett, D. I. (2002b). What competition? Response. *Trends in Cognitive Sciences,* 6(3): 119–119.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide. 2nd ed.* New York: Psychology Press.

Magnussen, S., & Greenlee, M. W. (1999). The psychophysics of perceptual memory. *Psychological Research,* 62(2–3):81–92.

Mewhort, D. J. K., Marchetti, F. M., Gurnsey, R., & Campbell, A. J. (1984). Information persistence: A dual-buffer model for initial visual processing. In: Bouma, H. & Bouwhuis D. G. (Eds.), *Attention and performance X: Control of language processes* (pp. 287–298). London: Lawrence Erlbaum Associates.

Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.

Pasternak, T., & Greenlee, M. W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience,* 6(2):97–107.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics,* 16(2):283–290.

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance,* 7(1):167–179.

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science,* 20(1):6–10.

Reeves, A. (2007). An analysis of visual masking, with a defense of 'Stopped Processing'. *Advances in Cognitive Psychology,* 3(1–2):57–65.

Schultz, D. W., & Eriksen, C. W. (1977). Do noise masks terminate target processing? *Memory and Cognition,* 5(1):90–96.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition,* 7(1–3):1–15.

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE,* 3(2):e1699.

Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology, 1,* 175:1–10.

Smith, W. S., Mollon, J. D., Bhardwaj, R., & Smithson, H. E. (2011). Is there brief temporal buffering of successive visual inputs? *Quarterly Journal of Experimental Psychology,* 64(4):767–791.

Smithson, H., & Mollon, J. (2006). Do masks terminate the icon? *Quarterly Journal of Experimental Psychology,* 59(1):150–160.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied,* 74(11):1–29.

Sperling, G. (1983). Why we need iconic memory. *Behavioral and Brain Sciences,* 6(1):37–39.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review,* 80(1):1–52.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition,* 15(3):181–198.

Watson, A. B., Borthwick, R., & Taylor, M. (1997). Image quality and entropy masking. *SPIE Proceedings,* 3016:2–12.